Lab 2 Practice: Introduction to Data

Background

The Centers for Disease Control (CDC) collects and monitors data on a multitude of health related issues. The ratio of male to female births can help to understand population growth trends and trends in infant mortality. Read *http://www.cdc.gov/nchs/data/nvsr/nvsr53/nvsr53_20.pdf* for more information about an analysis of sex ratios at birth in the United States (the reading not required to complete this lab!).

In the the DataCamp, you recreated some of the displays and preliminary analysis of Arbuthnot's baptism data. Your assignment involves repeating these steps, but for present day birth records in the United States.

Practice

As you are going through this practice:

- All R commands should go in R chunks!
- Be sure to label items with question numbers outside of the R chunk.
- Be sure to include commentary for questions that require written responses (#9 and #10) outside of R chunks.

1. Open the `R Reference Guide` from the `Lab Content` area of PolyLearn.
2. Open `RStudio` and then open a brand new `Rmarkdown` document by clicking on the green plus sign on the top left of RStudio. `Knit` this docment to see how the R code chunks work.
3. If you are using the public computers in the lab, you likely do not need to do this step. Otherwise, locate the `Packages` tab in the lower right hand pane of RStudio, and click on `Install`. One at a time, install both the `readr` and `mosaic` packages. You should only have to execute this on your personal machine this **one time** for the whole quarter.
4. Delete everything in the Rmarkdown document. Identify the `Rmarkdown` tab of your `R Reference Guide`. Copy and paste the three code chunks (header, set up chunk, and import data code chunk) into your Rmarkdown document.
5. Identify the `Lab Data Sets` folder on PolyLearn and download the `present` data set to a location on your computer (i.e., desktop, STAT 217 folder). Follow the steps in the `Importing` tab of the `R Reference Guide` to import the `present` data set and save your import code in an R chunk.
6. Use an R command to identify the variable (column) names in this data set.
7. Use an R command to identify the dimensions of the data frame. (*How many observations and variables are there?*)
8. Use an R command to identify the years that are included in this data set.
9. In the DataCamp class you observed from the `arbuthnot` data set that boys were born in greater proportion than girls in London from 1629-1710. Now examine the `present` data set - does the relationship hold true in the US from 1940-2002? Why or why not? (*Hint: create a new variable that represents the proportion of boy births.*)
10. Make a plot that displays the proportion of boys over time. Use the argument `type = "l"` (note that the character in quotes is an "el" and not a "one") in the `plot` command to change the plotting type from points to a line. Do you see any trends?
11. Re-do the plot with a more informative title, x-axis label, and y-axis label.
12. Submit this lab assignment as an **html** compiled from Rmarkdown. Make sure names of **all** present lab group members who contributed to this lab assignment are listed on the html file.